Polar Modelling And Segmentation Of Genomic Microarray Spots Using Mathematical Morphology

Jesus Angulo

Image Anal Stereol 2008;27:107-124
Original Research Paper

POLAR MODELLING AND SEGMENTATION OF GENOMIC MICROARRAY SPOTS USING MATHEMATICAL MORPHOLOGY

JESÚS ANGULO

Centre de Morphologie Mathématique, Mines de Paris, 35 rue Saint Honoré, 77300 Fontainebleau, France
e-mail: jesus.angulo@ensmp.fr
(Accepted May 23, 2008)

ABSTRACT

Robust image analysis of spots in microarrays (quality control + spot segmentation + quantification) is a requirement for automated software which is of fundamental importance for a high-throughput analysis of genomics microarray-based data. This paper deals with the development of model-based image processing algorithms for qualifying/segmenting/quantifying adaptively each spot according to its morphology. A series of morphological models for spot intensities are introduced. The spot typologies represent most of the possible qualitative cases identified from a large database (different routines, techniques, etc.). Then, based on these spot models, a classification framework has been developed. The spot feature extraction and classification (without segmenting) is based on converting the spot image to polar coordinates and, after computing the radial/angular projections, the calculation of granulometric curves and derived parameters from these projections. Spot contour segmentation can also be solved by working in polar coordinates, calculating the up/down minimal path, which is easily obtained with the generalized distance function. With this model-based technique, the segmentation can be regularised by controlling different elements of the algorithm. According to the spot typology (*e.g.*, doughnut-like or egg-like spots), several minimal paths can be computed to obtain a multi-region segmentation. Moreover, this segmentation is more robust and sensible to weak spots, improving the previous approaches.

Keywords: genomic microarray image, mathematical morphology, polar coordinates, shortest path segmentation, spot modelling, spot segmentation.

INTRODUCTION

DNA microarrays are an experimental bio-technology of growing importance in identifying sequences in genomes (genotyping experiments), in quantifying the presence (comparative genomic hybridization experiments) and expression levels (transcript experiment) of genes. The method basically consists in the detection and/or quantification of the hybridization signal of a DNA or RNA sample on an array of thousands of known oligonucleotide sequences (probes) that are printed as spots on a support (Brown and Botstein, 1999; Schena, 2003).

Spot finding and signal intensity determination are performed with the help of image analysis software. Recently it has been shown that segmentation methods can significantly influence microarray data precision (Ahmed *et al.*, 2004). Successful work on spot location and segmentation has already been done during the last years (Chen *et al.*, 1997; Steinfath *et al.*, 2001; Bozinov and Rahnenführer J, 2002; Yang *et al.*, 2002; Demirkaya *et al.*, 2005; Gottardo *et al.*, 2006). A comparative evaluation of performance can be found in (Lehmussola *et al.*, 2006). We have previously proposed an automatic spot segmentation based on advanced morphological operators (Angulo and Serra, 2003). This inner marker (spot center) plus outer marker (bounding box from rectangular grid) watershed-based segmentation yields satisfactory results for "normal" spots. However, it is observed, on the one hand, segmentation problems for low intensity spots or for spots on strong noisy background; and on the other hand, difficulties to define a right segmentation/quantification for structured spots (*e.g.*, doughnut-like and egg-like spots). In addition, several typologies of abnormal or irregular spots can be related to different problems of preparation of microarrays and consequently, a qualitative automatic evaluation of spots can be of help for flagging the suspect spots, a necessary step for data analysis.

This manuscript is an extended version of the conference paper presented in the *XII International Conference in Stereology (ICSXII)* (Angulo, 2007), held in Saint-Etienne (France) in August 2007. It is organised into two main parts and it deals with the development of model-based image processing algorithms for qualifying, segmenting and quantifying adaptively each spot according to its morphology. In the first part, we focus on the morphological modelling and automated classification of spots according to different typologies. Several models have been suggested for spot intensity distribution, including

special statistical models: a stochastic/geometric model (Balagurunathan and Dougherty, 2002), a scaled bivariate Gaussian density function (Steinfath *et al.*, 2001), a difference of two Gaussian densities or a cylinder (Wierling *et al.*, 2002), a polynomial-hyperbolic model (Ekstrom *et al.*, 2004), linear models based on PCA (Glasbey and Khondoker *et al.*, 2005). These models are typically used for image simulation or for fitting model parameters. We prefer here to propose a morphological model with spot typologies which represent most of the possible qualitative cases identified from a large database (different routines, techniques, etc.). Then, based on these spot models, a classification framework has been developed. The spot feature extraction and classification (without segmenting) is based on converting the spot image to polar coordinates, and after computing the radial/angular projections, calculating granulometric curves and derived parameters from the projections.

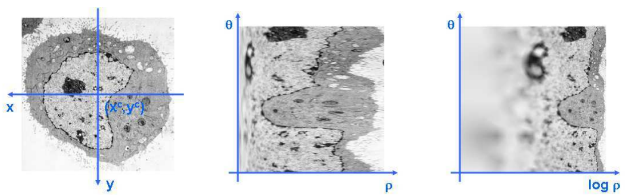
Furthermore, spot segmentation can also be approached in a more flexible and understandable way when working in polar coordinates. But the same weaknesses of the watershed on the low or noisy gradients are still underlying. The spot contour in polar coordinates is equivalent to calculating the left/right markers watershed-based transformation. This well-posed problem of segmentation can be also solved by calculating the up/down minimal path (easily obtained with the generalized distance function). The aim of the second part of the paper is just to introduce an innovative model-based spot segmentation according to this paradigm, where the type of segmentation is adapted to the spot typology. Several issues must be addressed, mainly the way for filtering the image on which the distance is computed and the manner to obtain a closed segmentation (circular shortest path). The shortest path segmentation can be regularised by controlling different elements of the algorithm. The segmentation of microarray spots in polar coordinates has also been addressed by (Appleton and Talbot, 2005), as an example of application of globally optimal geodesic active contours, but without considering the different typologies of spots. Another recent work has proposed a model-based spot segmentation by means of clustering algorithms (Li *et al.*, 2005).

The rest of the paper is organised as follows. In the second section we fix the notation and we give a reminder on image (log-)polar transformation. The third section introduces the image models for spots in microarray images. Then, in the fourth section the classification framework for polar-based spot classification is presented. In the fifth section the algorithm for computing the generalized distance global minimal paths is reminded. The sixth section introduces the polar-based spot segmentation by global minimal paths according to the spot typology. In the seventh sectionthe results obtained from a deep empirical study are discussed. Finally, the conclusions and perspectives are given in the eighth section. The paper is completed with two appendix sections. Appendix A provides some additional elements about the spot classification algorithm. Appendix B presents the algorithm to optimally compute the spot center, which is required for spot polar transformation.

NOTATION AND BASIC DEFINITIONS

In the framework of digital grids, a grey tone image associated to a scanned microarray can be represented by a function $f : E \rightarrow \mathscr{T} = \{t_{\min}, t_{\min}+1, \cdots, t_{\max}\}$, where E is a discrete space ($E \subset \mathbb{Z}^2$), domain of definition of the function f, and $\mathscr{T}$ is an ordered set of discrete grey-levels, *i.e.*, a subset of $\mathbb{Z}$. Typically, $t_{\min} = 0$ and $t_{\max} = 2^{16} - 1 = 65535$ for a 16-bits image file. $f(\mathbf{x})$ is the intensity value of the image at point $\mathbf{x} = (x,y)$.

The spots are structures placed regularly on the microarray image. Let the image zone $Z_i \subset E$ be defined as the influence cell (or bounding box region since the spots are usually placed in an orthogonal array structure) around spot i, *i.e.*, pixels of the zone where their distance to the center of spot i is lower than the distance to the other spot centers. Ideally, we can suppose that $Z_i \cap Z_j = \emptyset$, $\forall i, j \backslash i \neq j$ (*i.e.*, overlapping between neighbouring spots is impossible). The image signal intensity in the cell associated to the spot i at pixel position $\mathbf{x}$ is denoted by $f_i : Z_i \rightarrow \mathscr{T}$, where obviously $f_i(\mathbf{x}) = f(\mathbf{x})$, that is, function f_i is a restriction of function f to the set of support Z_i. In order to consider individually each spot but establishing spot models, we refer by $s_i(\mathbf{x} - \mathbf{x}_i^c) = f_i(\mathbf{x})$ function $s_i(\mathbf{y})$, $\mathbf{y} \in E$, translated at $\mathbf{x}_i^c$, the central point of spot i.



$f(x,y)$ $\quad f^{\circ}(\rho,\theta)$ $\quad f^{\circ}(\log\rho,\theta)$

Fig. 1. *Examples of polar and log-polar transformation of a grey-level image.*

The polar transformation converts the Cartesian image function $f(x,y) : E \rightarrow \mathscr{T}$ into another polar image function $f^{\circ}(\rho,\theta) : E_{\rho,\theta} \rightarrow \mathscr{T}$, where the angular coordinates are placed on the vertical axis and the radial coordinates are placed on the horizontal one. More precisely, with respect to a central point $\mathbf{x}^c = (x^c, y^c)$, we have:

$$\rho = \sqrt{(x-x^c)^2 + (y-y^c)^2}, \quad 0 \leq \rho \leq R\,, \quad (1)$$

and

$$\theta = \arctan\left(\frac{y-y^c}{x-x^c}\right), \quad 0 \leq \theta < 2\pi\,. \quad (2)$$

To determine the angular coordinate θ in the polar representation, it must be limited to an interval of size 2π. Conventional choices for such an interval are $[0,2\pi)$ and $(-\pi/2,\pi/2]$. To obtain θ in the interval $[0,2\pi)$, the following algorithm may be used:

$$\theta = \begin{cases} \arctan(\hat{y}/\hat{x}) & \text{if } \hat{x} > 0 \text{ and } \hat{y} \geq 0 \\ \arctan(\hat{y}/\hat{x}) + 2\pi & \text{if } \hat{x} > 0 \text{ and } \hat{y} < 0 \\ \arctan(\hat{y}/\hat{x}) + \pi & \text{if } \hat{x} < 0 \\ \pi/2 & \text{if } \hat{x} = 0 \text{ and } \hat{y} > 0 \\ 3\pi/2 & \text{if } \hat{x} = 0 \text{ and } \hat{y} < 0 \\ \text{undefined} & \text{if } \hat{x} = 0 \text{ and } \hat{y} = 0 \end{cases}$$

where $\hat{x} = x - x^c$ and $\hat{y} = y - y^c$. Now, the space support is $E_{\rho,\theta}$, $(\rho,\theta) \in (\mathbb{Z} \times \mathbb{Z}_p)$ (discrete period of p pixels equivalent to 2π). A relation is established where the points at the top of the image ($\theta = 0$) are neighbors to the ones at the bottom ($\theta = p-1$).

In many computer vision problems, the radial coordinate is replaced by the logarithm of ρ, named log-polar representation. The main advantage of the log-polar coordinates with respect to the polar coordinates is the fact that scale changes in the Cartesian image become horizontal shifts in the transformed image. In both polar and log-polar representations, rotations in the Cartesian image become vertical cyclic (*i.e.*, periodic) shifts in the transformed space. The application of morphological operators to images in (log-)polar coordinates has been recently studied by Luengo-Oroz *et al.* (2005). In Fig. 1 a comparison of the polar and the log-polar transformations of an image is given. We have chosen for the purposes of spot modelling and segmentation to work on the polar representation, which leads to a better resolution near the spot center since this resolution is needed for analysing structured spots. In addition, the log-polar requires a resampling of the Cartesian grid to improve the structure resolution close to the center.

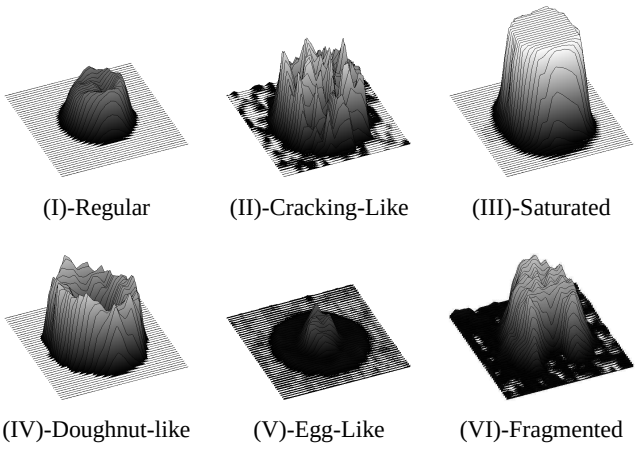


Fig. 2. *Examples of spot typologies.*

MODELS FOR SPOTS IN MICROARRAY IMAGES

Based on empirical observations of spots, we consider that the image intensity distribution for any spot i is given by the following expression:

$$f_i(\mathbf{x}) = a_i s_i(\mathbf{x} - \mathbf{x}_i^c) + n_i(\mathbf{x})\,, \quad (3)$$

where $s_i(\mathbf{y})$ corresponds to the morphological shape distribution for spot i. It is assumed for our purposes of classification and segmentation that s_i is represented by a cylindrical model. More precisely, a_i is the height of the "cylindrical" peak for spot i, $\mathbf{x}_i^c$ are the coordinates of the center position of the peak for spot i, and $n_i(\mathbf{x})$ is a function that describes the image noise.

Background noise

Two different sources of background noise can be distinguished:

$$n_i(\mathbf{x}) = n^g(\mathbf{x}) + n_i^l(\mathbf{x})\,. \quad (4)$$

$n^g(\mathbf{x})$ is the global background at point $\mathbf{x}$. This function can be typically described by a randomly Gaussian distributed noise for the whole image, *i.e.*, $n^g \sim N(\mu_n, \sigma_n^2)$. This part of the noise can be considered as associated to the acquisition system (photon-electronic scanner, CCD camera, etc.)

$n_i^l(\mathbf{x})$ is the local background noise (regionalised variable). It can be associated to different local phenomena: inhomogenous illumination, artefacts and inhomogeneities on the surface of support, errors in the preparation, etc.

Morphological spot typologies

The intensity distribution for spot i is a cylindrical peak with a variable radius and height:

$$s_i(\mathbf{y}) = r_i(\theta) t_i(\mathbf{y}) \; . \tag{5}$$

$r_i(\theta)$ is a "shape" function in polar coordinates describing the contour of spot i. It defines a closed boundary such that

$$s_i(\mathbf{y}) = \begin{cases} t_i(\mathbf{y}) & \text{if} \quad \|\mathbf{x} - \mathbf{x}_i^c\| \leq r_i(\theta) \; , \\ 0 & \text{if} \quad \|\mathbf{x} - \mathbf{x}_k^c\| > r_i(\theta) \; . \end{cases} \tag{6}$$

$t_i(\mathbf{y})$ is a "texture" function, that is, a spatial variable (more or less regular) function of intensity. Note that this structural variation of intensity at each point of the spot (biochemistry, hybridisation, washing and fixing, etc.), is different from the background noise.

According to the particular distributions of $r_i(\theta)$ and $t_i(\mathbf{y})$ in this model, it is possible to identify six main typologies of spots, see the examples of Fig. 2.

(I) Regular spot: In the case of a typical regular spot, the DNA material deposition on the spot is considered to be circular with an homogenous intensity distribution. The radius can be modeled by a normal distribution having mean μ_r and variance σ_r^2: $r \sim N(\mu_r, \sigma_r^2)$. Typically, the radius mean is random over a small range within the array and it can be considered as an uniform distribution, $\sigma_r \sim U(r_{\min}, r_{\max})$. The global variation of intensity, $a_i t_i(\mathbf{y})$ can be modeled as a normal distribution function, where the texture is a normal distribution with mean $\mu_t = 1$ and variance σ_t^2: $t \sim N(1, \sigma_t^2)$. Coefficient a_i is considered as the ground truth expression signal, modeled as another uniform distribution, $a \sim U(t_{\min}, t_{\max})$.

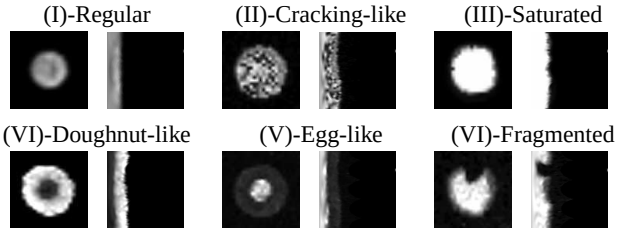


Fig. 3. *Examples of spot typologies in Cartesian and polar coordinates (in image pairs, left images are depicted in Cartesian and right in polar coordinates).*

(II) Cracking-like spot: The spot has an aspect of cracked or ripped intensity, *i.e.*, some dark tortuous lines or strips cross the spot surface. These zones typically result in low intensities levels. The radius shape function for $r_i(\theta)$ has the same normal distribution as for a typical spot. The texture function can be given by the equation $t_i(\mathbf{y}) = \widetilde{t}_i(\mathbf{y}) - \chi_i(\mathbf{y})$, where $\widetilde{t}_i(\mathbf{y})$ has the same model as the typical spot and where the cracking function $\chi_i(\mathbf{y}) > 0$ if $\mathbf{y} \in$ Crack Zone. The distribution of $\chi_i(\mathbf{y})$, the morphology of the strips (number, length, etc.) and spatial position are difficult to be modeled but typically, the strip thickness is significantly smaller than spot radius r.

(III) Saturated spot: The fluorescence saturated spots are characterised by a saturated intensity, *i.e.*, $a_i = t_{\max}$, with no variation of texture, *i.e.*, $t_i(\mathbf{y}) = 1$, and a regular shape of the contour, *i.e.*, $r_i(\theta)$ has the same normal distribution as for a typical spot.

(IV) Doughnut-like spot: The spot presents a circular "hole" in its center. The intensity distribution is the combination of two radial-defined texture functions:

$$t_i(\mathbf{y}) = \begin{cases} t_i^{\text{low}}(\mathbf{y}) & \text{if} \quad \|\mathbf{x} - \mathbf{x}_i^c\| < r_i^{\text{in}}(\theta) \; , \\ t_i^{\text{high}}(\mathbf{y}) & \text{if} \quad r_i^{\text{in}}(\theta) \leq \|\mathbf{x} - \mathbf{x}_i^c\| \leq r_i^{\text{ou}}(\theta) \; . \end{cases} \tag{7}$$

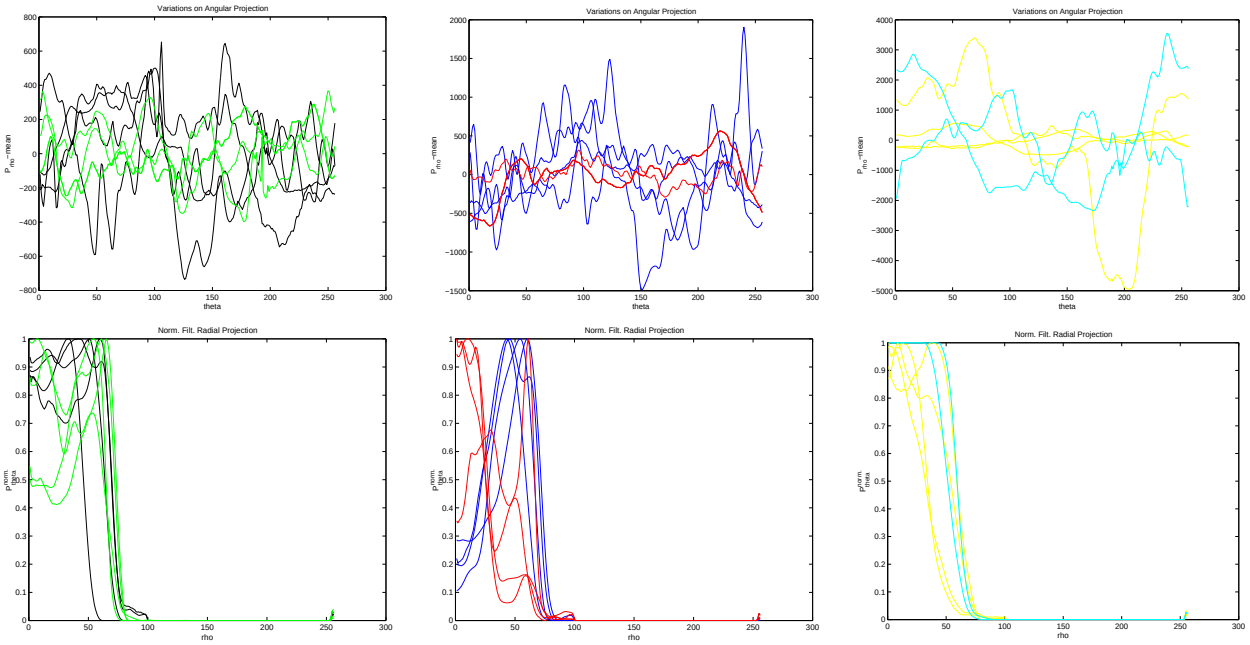
where $t_i^{\text{low}}(\mathbf{y})$ and $t_i^{\text{high}}(\mathbf{y})$ are the texture functions associated to the central part and to the peripheral part respectively; and $r_i^{\text{in}}(\theta)$ and $r_i^{\text{ou}}(\theta)$ are the radius functions of the center and of the spot contour respectively. We suppose that the inner and outer radius shape functions $r_i^{\text{in}}(\theta)$ and $r_i^{\text{ou}}(\theta)$ have the same normal distribution as for a typical spot (with mean $\mu_{r^{\text{in}}}$ and $\mu_{r^{\text{ou}}}$). In a similar way, the texture functions $t_i^{\text{low}}(\mathbf{y})$ and $t_i^{\text{high}}(\mathbf{y})$ have a normal distribution. Moreover, usually, the mean for $t_i^{\text{low}}(\mathbf{y})$ tends to 0 and the mean for $t_i^{\text{high}}(\mathbf{y})$ tends to 1.

We can consider also the **ring-like spot** as a degenerated case of the doughnut-like spot such that $(\mu_{r^{\text{ou}}} - \mu_{r^{\text{in}}}) \leq \delta$ (being relatively small), that is, the central hole is very large with δ significantly smaller than $\mu_{r^{\text{ou}}}$.

(V) Egg-like spot: Dual to the precedent, this spot has also two superposed intensity levels. More precisely, a circular peak of intensity t_i^{high} centered at position $\mathbf{x}_i^{ci}$ (but not necessarily with $\mathbf{x}_i^{ci} = \mathbf{x}_y^c$) which is added to a pedestal of intensity t_i^{low}, *i.e.*,

$$t_i(\mathbf{y}) = \begin{cases} t_i^{\text{high}}(\mathbf{y}) & \text{if} \quad \|\mathbf{x} - \mathbf{x}_i^{ci}\| < r_i^{\text{in}}(\theta) \; , \\ t_i^{\text{low}}(\mathbf{y}) & \text{if} \quad (r_i^{\text{in}}(\theta) \leq \|\mathbf{x} - \mathbf{x}_i^{ci}\|) \\ & \quad \text{and} \; (\|\mathbf{x} - \mathbf{x}_k^c\| \leq r_i^{\text{ou}}(\theta)) \; . \end{cases} \tag{8}$$

The inner and outer radius shape functions $r_i^{\text{in}}(\theta)$ and $r_i^{\text{ou}}(\theta)$, and the texture functions $t_i^{\text{low}}(\mathbf{x})$ and $t_i^{\text{high}}(\mathbf{x})$ have typically normal distributions, where here typically $\mu_{t^{\text{high}}}$ tends to 1 and $\mu_{t^{\text{low}}} > 0$.



In black, (I)-Regular
In green, (II)-Cracking-Like

In blue ,(IV)-Doughnut-like
In red, (V)-Egg-Like

In cyan, (III)-Saturated
In yellow, (VI)-Fragmented

Fig. 4. *Angular projections $P_\rho(\theta)(f_i^\circ)$ (top row) and radial projections $P_\theta(\rho)(f_i^\circ)$ (bottom row) for a selection of representative spots of each typology.*

In this case, we suppose that in the global variation of intensity $a_i t_i(\mathbf{x})$, the value of $a_i = 1$. The estimate of the mean or the median intensity of $t_i(\mathbf{x})$ cannot be adequate as a spot parameter. The same considerations are valid for the Doughnut-like spots.

(VI) Fragmented spot: A fragmented spot is characterised by a degenerated or irregular shape function $r_i(\theta)$, having also a size (surface area) lower than the typical spot within the array. The standard deviation σ_r is relatively important with respect to the mean. The texture function $t_i(\mathbf{y})$ can still be modeled as a normal distribution.

MODEL-BASED SPOT CLASSIFICATION

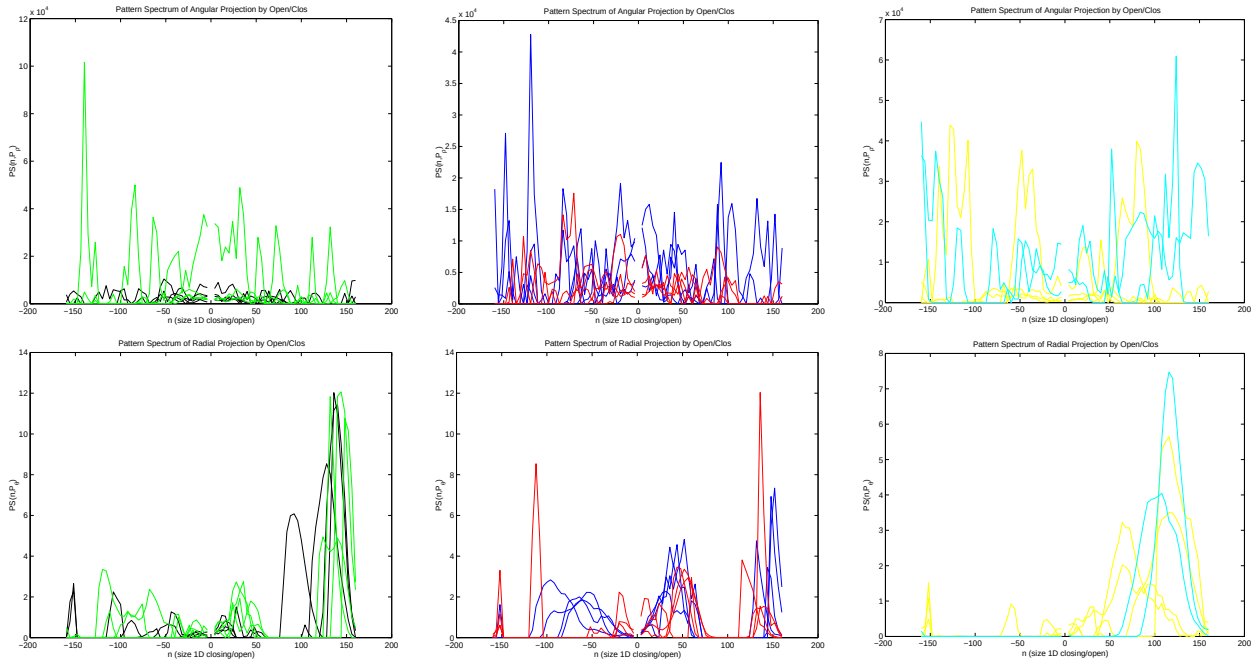
Based on the spot models introduced above, we have developed a classification framework for the different spot typologies. The algorithms for feature extraction and classification must be simple and fast: each spot should be individually processed and typical microarrays have thousands of spots. The parameters and the typology will be used to improve and to make the result of segmentation/quantification more robust.

Spots in polar coordinates: According to the models proposed, the polar representation seems to be appropriate to characterise the different spot distributions. Let $f_i^\circ(\rho,\theta)$ be the image polar representation of spot i. Fig. 3 gives an example of a spot for each typology. As pointed above, we have compared it with the log-polar representation and verified that it is more interesting to work on polar images for texture analysis. The "optimal" center point $\mathbf{x}_i^c$ for each spot is obtained by means of the algorithm presented in Appendix B of the paper.

Angular and radial projections: The horizontal and vertical projections of image $f_i^\circ(\rho,\theta)$ are then used to describe the spot structures: angular projection $P_\rho(\theta)(f_i^\circ) = \sum_{\rho=0}^{R} f_i^\circ(\rho,\theta)$ and radial projection $P_\theta(\rho)(f_i^\circ) = \sum_{\theta=0}^{p-1} f_i^\circ(\rho,\theta)$. Fig. 4 provides the projections $P_\rho(\theta)$ and $P_\theta(\rho)$ for a selection of spots from each typology.

From the analysis of $P_\rho(\theta)$ using Fourier descriptors or morphological parameters (Angulo, 2005), we state that its variation combines the contributions of the background and the spot, including the texture and the shape irregularities. Consequently $P_\rho(\theta)$ is a very poor descriptor to discriminate spot typologies. As we show below, $P_\theta(\rho)$ is more useful for spot classification.

Morphological filtering of $P_\theta(\rho)$: We start by extracting the background contribution using the top-hat transformation followed



In black, (I)-Regular
In green, (II)-Cracking-Like

In blue ,(IV)-Doughnut-like
In red, (V)-Egg-Like

In cyan, (III)-Saturated
In yellow, (VI)-Fragmented

Fig. 5. *Pattern spectra of angular projection, $PS(n_\theta, P_\rho(\theta))$, (top row) and pattern spectra of radial projection, $PS(n_\rho, P_\theta(\rho))$, (bottom row) for a selection of representative spots of each typology.*

by a normalisation, *i.e.*, $P^*_\theta(\rho) = P_\theta(\rho) - \gamma_n(P_\theta(\rho))$ and $\overline{P}_\theta(\rho) = P^*_\theta(\rho)/\max_{P^*_\rho(\theta)}$. The value $\sigma^\downarrow = \sum_{\rho=0}^{R} \gamma_n(P_\theta(\rho))/\sum_{\rho=0}^{R} P_\theta(\rho)$ gives an estimate of the regional background. Finally, a pre-filtering step is necessary in order to remove the insignificant extrema, *i.e.*, $\overline{P}^h_\theta(\rho) = \varphi^{rec}\left(\overline{P}_\theta(\rho)+h;\ \gamma^{rec}(\overline{P}_\theta(\rho)-h;\ \overline{P}_\theta(\rho))\right)$, where typically $h = 2\%$ to 5% of the maximum of $\overline{P}_\theta(\rho)$ (which is equal to 1 since it has been normalised). We can now compute several parameters from the processed curves $\overline{P}^h_\theta(\rho)$ such as: an approximation of spot radius, the value for $\rho = 0$, standard deviation, the percentage of points equal to 1, *etc.*, which allow detecting the main typologies. In Appendix A of the paper the precise definition of parameters and the corresponding values for a selection of representative spots of each typology are given.

Granulometric analysis of $P_\theta(\rho)$: Furthermore, the variation of $P_\theta(\rho)$ can be analysed by means of 1D granulometries or pattern spectra. A granulometry is a family of openings of increasing size $\{\gamma_n\}_{n\geq 0}$ and the pattern spectrum of f is the following mapping $PS_\gamma(f,n) = (m(\gamma_n(f)) - m(\gamma_{n+1}(f)))/m(f),\ n \geq 0$ and where $m(g)$ is the integral of g. A dual definition $PS_\varphi(f,-n)$ is associated to a family of closings and then both curves are represented together $\{-n,0,n\} \rightarrow PS(f,n) = \{PS_\varphi(f,-n),0,PS_\gamma(f,n)\}$. Note that the computation of these 1D openings/closings is very fast. In Fig. 5 the corresponding pattern spectra for the selection of spots are shown. The significant parameters computed from $PS(n_\rho, P_\theta(\rho))$ (see definitions and some examples in Appendix A) combined with those obtained directly from $P_\theta(\rho)$ allow a spot classification into the different typologies considered and without needing the spot segmentation. (More details in Angulo (2005)).

GENERALIZED DISTANCE GLOBAL MINIMAL PATH ALGORITHMS

Limitations of watershed transformation for detecting lines: According to the analysis by Vincent (1998), extracting a continuous track (="crest-line") going from the top to the bottom of the image by means of a constrained watershed, using as markers the right and left sides of the image, presents several limitations: (1) it fails when *SNR* (= sensitivity of watershed line to noise) is low; (2) the watershed between two markers A and B depends on the position of the saddle points (for all the paths joining A to B with minimal elevation, the highest pixels along those paths are the saddle points) between the markers, and their location is one

of the main factors determining the location of the line; (3) the criteria used to build the watershed are based on grey levels, and the length of watershed lines is irrelevant. Length constraints can be introduced in the segmentation by using global minimal path algorithms. This approach is also useful to detect "disconnected" crest-line between two markers.

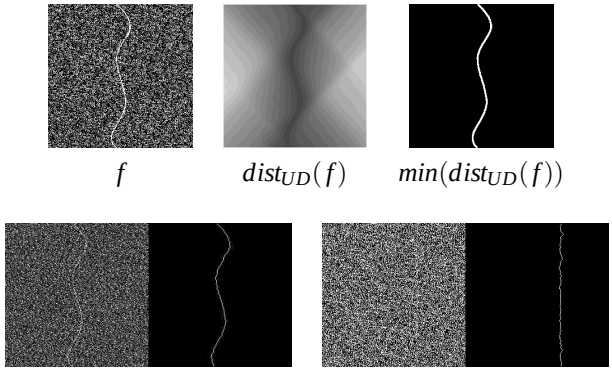


Fig. 6. *Top, generalised distance function and global minimal paths. Bottom, two examples of GMP detection in very noisy images.*

Generalised distance function, GDF: The algorithm is based on a modification of the classic two-pass sequential distance function algorithm of Rosenfeld and Pfaltz (1968) so that: (1) edge cost is taken into account; (2) raster and anti-raster scans are iterated until stability. Let us denote by $N^+(p)$ (resp., $N^-(p)$) the neighbors of pixel p scanned before p (resp., after p) in a raster scan, for a 8-connected grid (neighborhood graph). In this graph, to each edge between two neighboring pixels p and q of an image f one associates the cost value $C_f(p,q) = f(p) + f(q)$ (or any other monotonically increasing function, such as $\max(f(p), f(q))$ or $\min(f(p), f(q))$). More specifically, the algorithm of GDF to set X in image f proceeds as follows,

- Initialise result image d: $d(p) = 0$ if $p \in X$ and $d(p) = +\infty$ otherwise;
- Iterate until stability:
  - Scan image in raster order $\rightarrow$ For each pixel p, do: $d(p) \leftarrow min\{d(p), min\{d(q) + C_f(p,q), q \in N^+(p)\}\}$
  - Scan image in anti-raster order $\rightarrow$ For each pixel p, do: $d(p) \leftarrow min\{d(p), min\{d(q) + C_f(p,q), q \in N^-(p)\}\}$

Depending on the cost value considered, the algorithm typically converges in two or three iterations (relatively efficient).

Global minimal paths, GMP: Each path P in the 8-connect graph has an associated cost $C_f(P)$, equal to the sum of the cost of its successive edges. We can now define the distance $d_f(p,q)$ between two pixels p and q in the image f as: $d_f(p,q) = \min\{C_f(P), P \text{ path between } p \text{ and } q\}$.

For the simple problem of finding a path of minimal cost (or global minimal path, GMP) going from the top row U to the bottom row D of the image, we use the following result: a pixel p belongs to such a minimal path if and only if $d_f(p,U) + d_f(p,D) = d_f(U,D)$. This is the approach introduced by Vincent (1998). To extract such Up/Down GMP in image f, we can therefore proceed as follows:

- Compute GDF to set U in image f: for each pixel p, compute $d_f(p,U)$;
- Compute GDF to set D in image f: $d_f(p,D)$;
- Sum these two distance functions, $d_f(U,D)(p) = d_f(p,U) + d_f(p,D)$;
- Find $u_{\min}$, the minimal value of $d_f(U,D)$ and threshold the result in order to keep only the pixels which values in $d_f(U,D)$ are equal to $u_{\min}$.

Since the extracted minimal paths are preferentially located on dark pixels (*i.e.*, have low cost), the original image with the bright track must be inverted before computing the two generalised distance functions. From an algorithmic point of view, the problem is reduced to computing two grey-weighted generalised distance transforms. Fig. 6 shows some examples, illustrating the robustness against the noise.

To give priority to the "vertical" paths, the computation of the distance function is constrained for raster scan to the top-left, top-middle and top-right pixels of p in the neighborhood $N^+(p)$ (resp., bottom-left, bottom-middle and bottom-right for anti-raster scan $N^-(p)$). Another way to formulate it is to say that at any location along a track, according to the neighborhood graph used, it is assumed that the absolute value of the angle between the track and the vertical direction is less than or equal to 45°. It guarantees a certain smoothness to the extracted tracks. This segmentation can be interpreted in terms of an optimality criteria framework (Vincent, 1998): (1) the pixel values along the track (to maximise), (2) the length of the track (to minimise), (3) the raggedness of the track (to minimise).

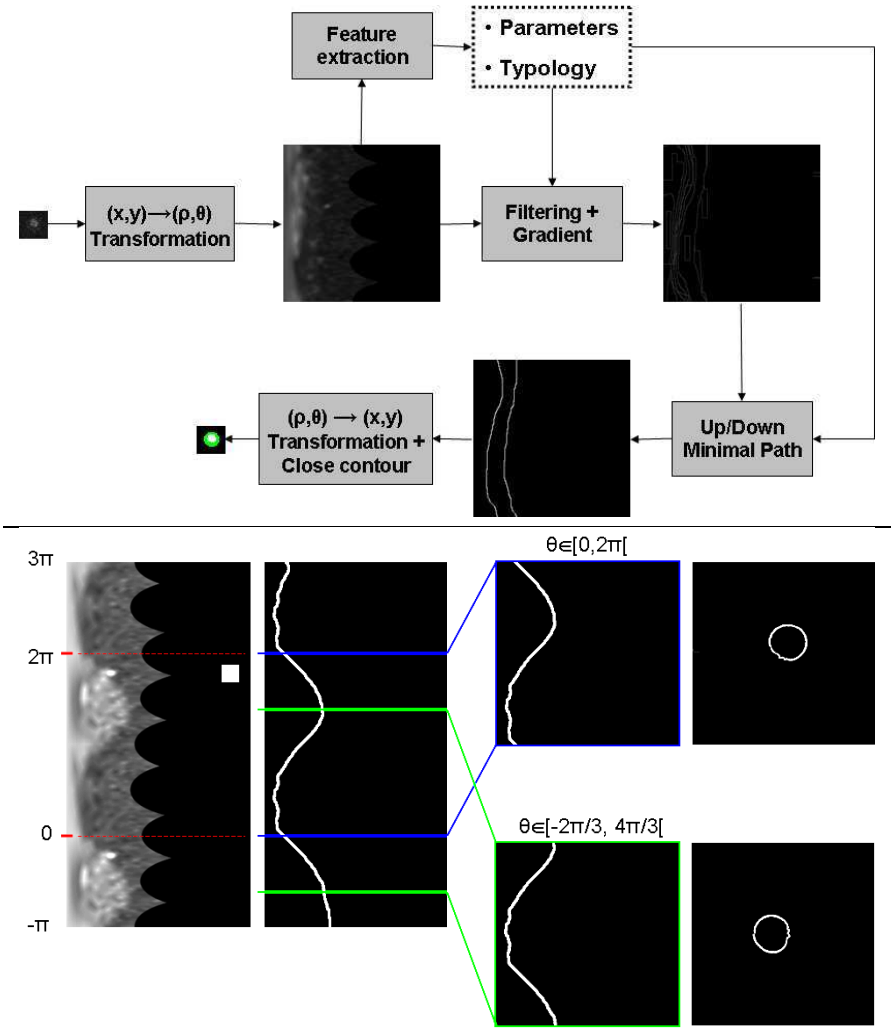


Fig. 7. *Top, flowchart of algorithm for model-based spot segmentation in polar coordinates. Bottom, selection of a period of cyclic minimal path to define close contour.*

MODEL-BASED SPOT SEGMENTATION BY MIN. PATHS

Starting from the gradient of a filtered version of the spot in polar coordinates, the aim now is to segment its contour using the GMP technique. To achieve a robust algorithm several issues must be considered in detail (see the diagram of Fig. 7).

Interpolation: The spots are small image structures, typically their diameter is approximately equal to 15 pixels and their bounding boxes of size 25×25 pixels. In polar coordinates, the radial variation is consequently limited to around 7 pixels. These small magnitudes limit the possibilities to obtain a regularised or multi-level segmentation of the spots by using Up/Down GMPs. Let $f_i(\mathbf{x})$ be the original spot sub-image, we propose to interpolate it by using a bi-linear schema to increase the size of structures to be segmented, $f_i^{\uparrow k}(\mathbf{x})$ (a factor $k = 4$ constitutes a typical value considered in all our examples). The Cartesian-to-polar conversion is then computed from this image, followed by the Up/Down GMP and the inverse conversion. The advantages of this interpolation are:(1) to increase the accuracy of segmentation; (2) to allow a contour regularisation by a larger choice for the different sizes of filtering; (3) the spot region could be segmented into several regions using multiple GMPs (obviously, the closed contour must be decimated by the same factor k in order to obtain the original spot size); (4) spot feature extraction and spot classification is also obtained from this enlarged representation.

Circular minimal path to close contour: In order to obtain a closed contour for the spot region, we must impose a circular minimal path, *i.e.*, in polar coordinates and with the Up/Down GMP, the initial radial value ρ_{up} (for $\theta = 2\pi$) and the final one ρ_{down} ($\theta = 0$) are equal. Several algorithms have been

proposed in the literature to calculate circular minimal paths, relatively sophisticated and solved by dynamic programming (multiple search algorithm, branch and bound algorithm, etc.) (Sun and Pallottino, 2003). We propose to apply a simpler algorithm to allow using GMP approach to define closed spot contours.

The original polar images $[0,2\pi[$ can be cycled, extending the image along its angular direction by adding the top part of the image on the bottom and the bottom part on the top, and consequently repeating another period of the image. When the Up/Down GMP is applied to this cycled image, the continuity provided by the added cycle yields almost always a circular path. In fact, even if $\rho_{up} \neq \rho_{down}$, but $|\rho_{up} - \rho_{down}| \leq \Delta_\rho$ (Δ_ρ being a small value, typically 2 to 5 pixels), the contour can be "closed" applying previously a dilation of size Δ_ρ before computing the transformation to Cartesian coordinates. Moreover, the cycled image allows to select different periods of the minimal path to find a circular minimal path or at least the minimal path with the lowest Δ_ρ. In practice, the translation along the angular axis θ in polar coordinates involves a rotation in Cartesian coordinates, *i.e.*, if the selected period of $\theta \equiv [0+\alpha, 2\pi+\alpha[$ the image of the closed contour should be rotated α radians. To avoid the vagueness due to the rotation, we usually consider five simple cases ($\alpha = 0$, $\pi/2$, π, $-\pi/2$ and $-\pi$) and we choose the α which has lower Δ_ρ (see example in Fig. 7).

Filtering and gradient in polar coordinates: As we have shown, the polar image $f_i^\circ(\rho,\theta)$ is cycled to ensure the periodicity of the angular coordinate. The polar image filtering (*i.e.*, type and sizes of filters) is a critical step in order to achieve a robust segmentation method.

An anisotropic effect in polar coordinates is obtained by applying two separable directional filters (unidimensional filtering) in the angular and radial coordinates. Usually, for the polar image of spots, the vertical (according to the angular coordinate) filtering has a size n_θ which is notably higher than the size n_ρ of horizontal filtering (radial direction). We have compared three different families of filters: Gaussian diffusion, morphological operators (opening/closing + levelling) and sliding average. In fact, the average filter is the simplest and fastest approach which simplifies the structure in such a way that the GMP corresponds to the main spot contour. It seems that the sizes $n_\rho = 16$, $n_\theta = 48$ ($\simeq \pi/3$) yield a satisfactory trade-off for this spot whose diameter is approx. equal to 7 pixels ($7 \times 4 = 28$ pixels in the interpolated version). If the adequate vertical size of filtering can be considered as independent from the spot diameter, the choice for the horizontal one well-adapted to one spot is obviously associated to an estimate of its radius, obtained from the radial projection (see previous section). Concerning the gradient, the external gradient is always applied, $g^+(f(\mathbf{x})) = \delta_1(f(\mathbf{x})) - f(\mathbf{x})$.

Spot typologies for segmentation: The *homogeneous spots* (regular or saturated) are easily segmented using the present approach. The *inhomogenous spots* (cracked or fragmented) need an estimate of the spot diameter and of the texture degree to adapt the size of the horizontal/vertical anisotropic filtering. In the case of *empty spot* (or absent spot), we propose to calculate also a GMP to segment the background and try to compute a parameter of intensity. These classes of spots only need one contour. The segmentation of *doughnut-like spots* (*i.e.*, presenting a hole) and *egg-like spots* (*i.e.*, with a peak of intensity) needs the computation of a multiple contour, *i.e.*, multiple minimal path.

Several alternatives can be applied for the spot segmentation in two or more regions. From a mathematical morphology viewpoint, this involves filtering the spot, removing the hole/peak, and therefore enhancing its main contour. In order to do that, we use the "close-holes" operator. This operator fills all holes in an image f that does not touch the image boundary f_∂ (used as a marker) and therefore provides a parameter free approach to detect holes in an image: $\psi^{\text{ch}}(f) = [\delta_{f^c}^{rec}(f_\partial)]^c$, where $\delta_g^{rec}(m)$ is the geodesic reconstruction of the marker image m within the reference image g. For a binary image, the definition of grains and holes is clear; for the case of grey level images, a "hole" is defined as a set of connected points surrounded by connected components of value strictly greater than the hole values. This operator is a morphological closing and therefore removes the dark structures (valleys of intensity). A dual version of this operator, $\psi^{\text{ch-dual}}(f) = [\psi^{\text{ch}}(f^c)]^c$, allows the definition of a dual close-holes operator to remove the peaks of intensity.

We can also work on the residues of these morphological operators. That is, to be able to segment, on the one hand, the spot without hole or grain and on the other hand, the hole and the grain. The final algorithm proposed is based just on working on three different images on which we compute eventually, and according to the typology, up to three Up/Down GMP. This algorithm can be summarised as follows. Let f_{spot} be the original spot image, to apply the following steps according to the spot typology:

1. Obtain the hole image, *i.e.*, $f_{\text{spot}}^{\text{ch}} = \psi^{\text{ch}}(f_{\text{spot}})$; $f_{\text{spot}}^{\text{hole}} = \psi^{\text{ch}}(f_{\text{spot}}) - f_{\text{spot}}$.

2. Obtain the enhanced spot image and the peak image, *i.e.*, $f_{\text{spot}}^{\text{ch-dual}} = \psi^{\text{ch-dual}}(f_{\text{spot}})$; $f_{\text{spot}}^{\text{peak}} = f_{\text{spot}} - \psi^{\text{ch-dual}}(f_{\text{spot}})$.
3. Compute the centroid for image f_{spot} (see Appendix B). Eventually, compute the centroid for corresponding hole or peak images.
4. Calculate main contour of spot by applying the algorithm Up/Down GMP to the enhanced spot image $f_{\text{spot}}^{\text{ch-dual}}$.
5. Calculate hole contour by applying the algorithm Up/Down GMP to $f_{\text{spot}}^{\text{hole}}$.
6. Calculate peak contour by applying the algorithm Up/Down GMP to $f_{\text{spot}}^{\text{peak}}$.

RESULTS AND DISCUSSION

The algorithms of spot classification and segmentation have been evaluated on various cDNA microarray images. This section summarises the most significant results obtained. Fig. 8 depicts two examples of spot segmentation using the present Polar GMP algorithm. As we can observe from these two typical blocks of microarray images, the contours obtained for the spots are very precise.

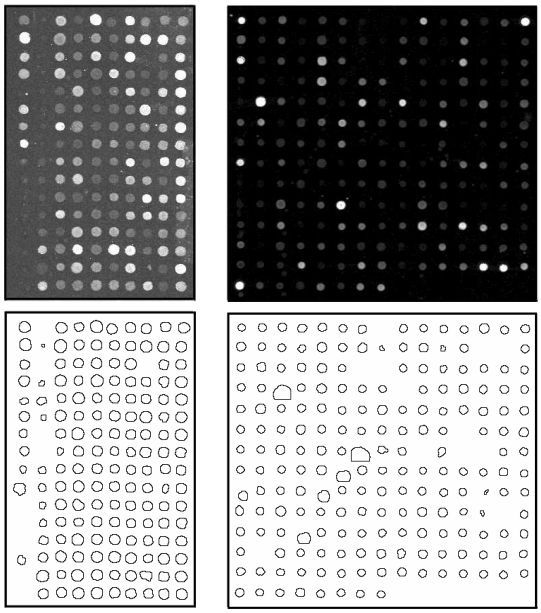


Fig. 8. *Two examples of spot segmentation using the present polar GMP algorithm: Top, original microarrays (for visualisation purposes, the intensity has been modified by a gamma function $\gamma = 2$). Bottom, contours of segmented spots.*

However, using these examples the potential of the approach cannot be appreciated since all the spots in both images are very regular. In fact, in the first image, only three spots are considered as fragmented spots and the rest are classified as regular spots. A criterion based on a minimal integral of intensity (> 50) after an area opening is used to consider an empty spot (as previously proposed in Angulo and Serra, 2003).

In Fig. 9 a more challenging example of spot classification and segmentation of a ("bad" quality) microarray image, including regular spots, doughnut-like spots and a majority of egg-like spots is given. To evaluate the performance of the present GMP approach in comparison with our previous method based on watershed segmentation (Angulo and Serra, 2003), we have chosen two replicated blocks (with the same DNA probes on each spot) from two different microarray images. In the Polar GMP segmentation layer image, the main contour of each spot appears in red, and the second contour is drawn in green for doughnut-like spots and in blue for egg-like spots. By a visual comparison of the segmentations, one observes that the results from the classical watershed-based approach seem satisfactory and quite coherent between both images. The results obtained from the Polar GMP approach are also coherent between both block images: most of the spots are classified as egg-like spots, and only four (in block 1) and two (in block 2) are classified as doughnut-like spots; and these structured spots are in most of cases correctly segmented in two regions. It seems that only the (quasi-)empty spots are segmented by a vague (or wrong) circular shortest path.

Nevertheless, apart from the visual analysis, a quantitative assessment of segmentation is needed to compare the results. The plots provided in Fig. 10 summarise the parameters computed from the segmented images of Fig. 9.

Starting from the size and shape of the contours, it is evident that the spot regions of the watershed-based segmentation are a bit more uniform between both blocks: average area of 351 ± 108 pixels for Block 1 and 328 ± 112 for Block 2, with an average error of area between equivalent spots of 67 ± 87; instead of 340 ± 118, 309 ± 170 and 130 ± 155 respectively for the main contour of the polar GMP-based segmentation.

However, the spot contours obtained with the new approach present a more regular form factor (defined as $perimeter^2/4\pi area$): 1.03 ± 0.20 for Block 1 and 1.02 ± 0.17 for Block 2 (the values for the watershed-based approach are respectively 1.09 ± 0.15 and 1.04 ± 0.10), which involves a good fit with a circular shape.

Table 1. *Statistical summary of relationships between the measured ratio of intensities Red/Green of both segmented Block 1 and Block 2 of Fig. 9 (see the text for more details).*

Ratio of integrals	Mean of error	Std.dev. of error	Coeff. of correlation
Watershed-based single contour	0.24	0.46	0.55
Polar GMP-based main contour	0.22	0.37	0.62
Polar GMP-based secondary contour	0.23	0.36	0.66
Polar GMP-based integrated contours	0.18	0.27	0.76
Ratio of medians	Mean of error	Std.dev. of error	Coeff. of correlation
Watershed-based single contour	0.34	0.52	0.54
Polar GMP-based main contour	0.06	0.11	0.62
Polar GMP-based secondary contour	0.08	0.11	0.66
Polar GMP-based integrated contours	0.11	0.08	0.77

But the most important value measured from the DNA microarray experiments is the ratio between the intensities of the red and the green images in each spot. Two main parameters can be defined as the "intensity" of the spot in each colour: the integral of image intensities inside the spot contour or the median (more robust than the mean) of intensities of the spot contour; and then, the ratio of integrals or the ratio of medians could be considered.

These two different ratios have been computed for the various segmentations. In Fig. 10 the corresponding values for the first column of spots are given and in Table 1 it is provided a statistical summary of relationships of two ratios between Block 1 and Block 2 for the various segmentations. The error is defined as the difference between the ratio of Block 1 and the same ratio of Block 1. We have compared the ratios associated to the single watershed-based contour with those associated to the main contour, the secondary contour and the average ratio of main and secondary contours (named ratio of integrated contour) for the polar GMP segmentation. It is evident that, for the three possible cases of the polar GMP segmentation, the ratio based on the median is more coherent between both Blocks. In the case of the watershed segmentation, the difference between both kinds of ratio is not significant. Clearly, for any of the three alternative contours from the new approach, the ratio of median intensities is more robust (lower errors between both Blocks) and more coherent (higher correlation between both Blocks) than the one of the single contour of the watershed-based approach.

The last point to discuss is the pertinence of having the structured spots segmented in two regions, and consequently two ratios of intensities describing each spot. From the analysis of results of the current example, we can state that (even if globally the integrated ratio, defined as the average of the ratio of main contour and the ratio of secondary contour, seems to lead to a better correlation) the most appropriate is to consider two separated ratios for the subsequent data analysis steps.

We consider that this empirical demonstration of better segmentation models and algorithms validates the contributions of this paper.

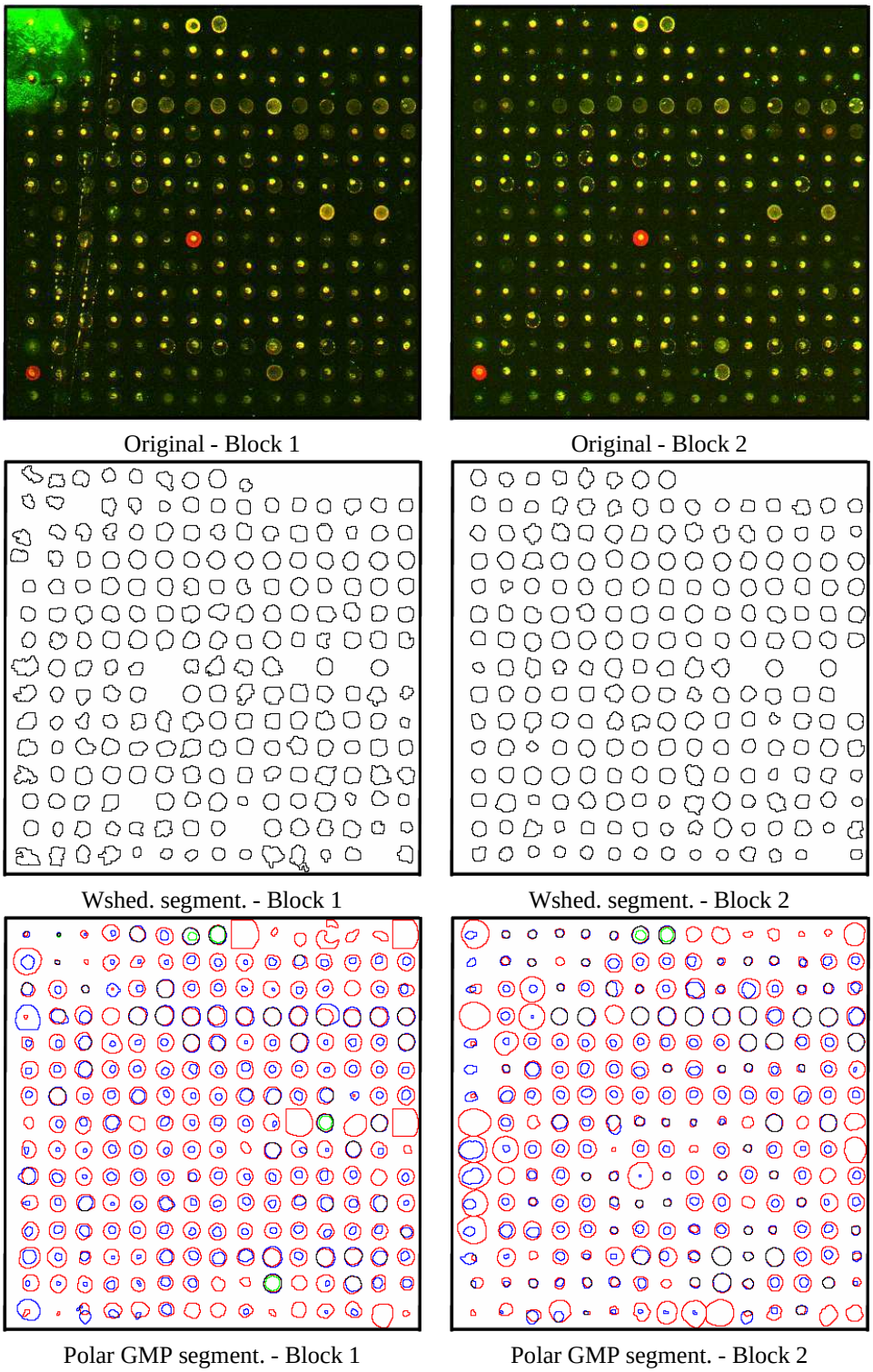


Fig. 9. *Comparison of spot segmentation on replicated blocks (the same DNA probes on each spot from two different microarray images): Top, original blocks of spots (for visualisation purposes, the intensity has been multiplied by* 10*); middle left, segmentation using classical watershed-based approach (by Angulo and Serra (2003)) of Block 1, bottom left, segmented spots according to the the present polar GMP algorithms (main contour in red, "peak contour" in blue and "hole contour" in green); right, idem. for Block 2.*

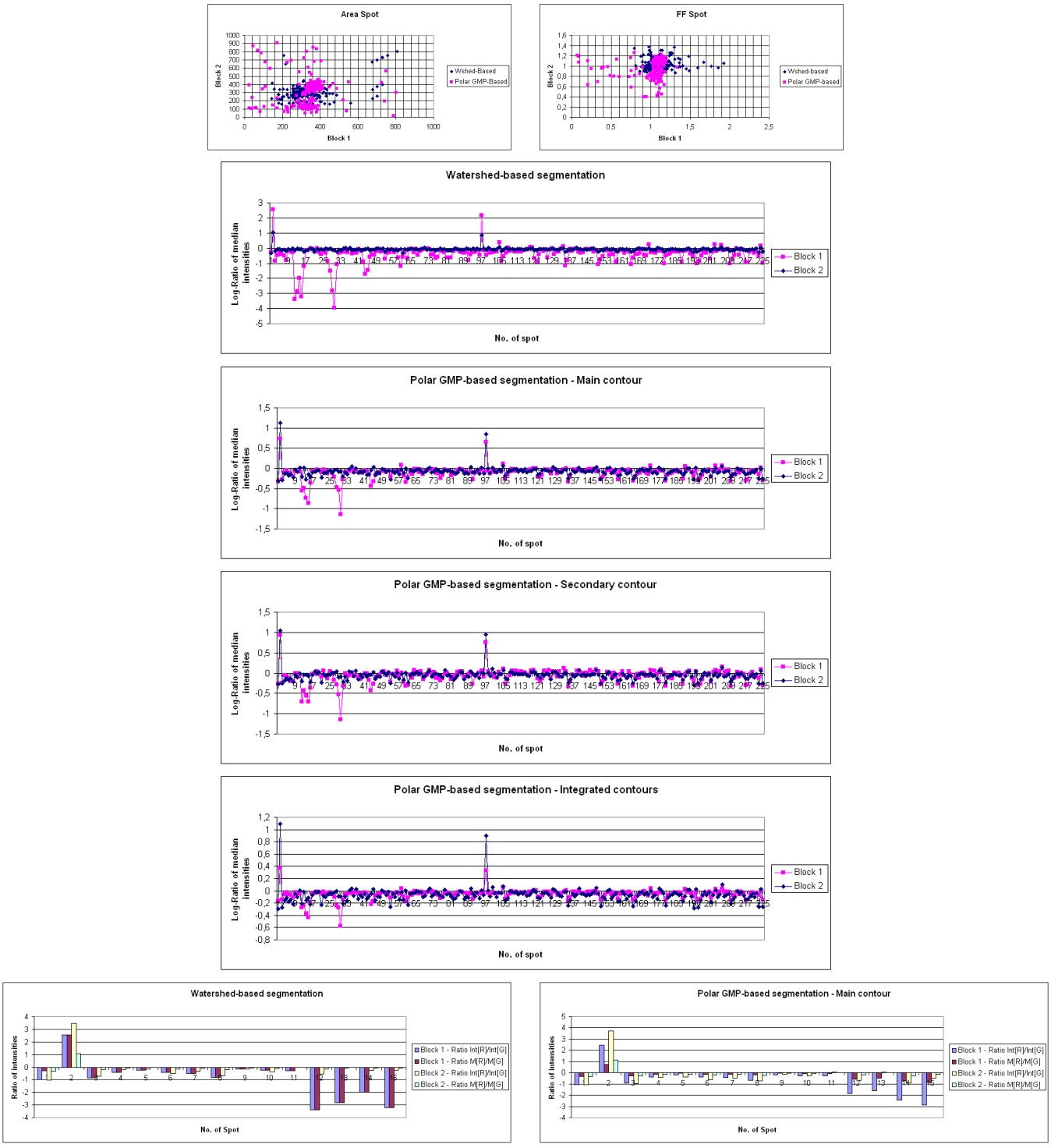


Fig. 10. *Summary of quantitative comparison of watershed-based vs. polar GMP-based segmentation of block microarray images of Fig. 9.*

CONCLUSION

Apart from supporting spot segmentation, morphological spot modelling allows calculating quality control parameters to eventually detect the preparation accidents. The proposed models can also be used to define distances between spots and spot kernels for image-based machine learning and classifying algorithms.

The results of model-based spot segmentation are satisfactory in terms of sensitivity and robustness and they improve the previous approaches, allowing an automatic adaptation to all spot morphologies and image qualities.

The algorithm comprises a pipeline of several complex steps and the spot regions should be classified and segmented one by one; consequently, the execution is not as fast as the watershed-based approach which simultaneously segments all spots (*e.g.*, a few *secs* for a typical block). In our current implementation, the C++ code of the present approach, running in a typical Laptop (Intel® Core™2 CPU 1.99 GHz and 1.99 GB RAM) takes approximately 1 sec to process each spot; consequently, between 2 and 3 min for a typical block. However, it is possible for a particular application to optimise various steps of the algorithm in order to accelerate the time execution.

An additional control step could be included in the algorithm in order to evaluate the pertinence of the minimal paths extracted (using, for instance, the value of the gradient along the path, the contrast between the regions separated by the path, etc.) which will be useful for the data post-processing and quantification.

Classically, the "intensity parameter" of the spot is given by computing the integral or the mean/median (and the variance) of the grey-level image points inside the spot region. In this new approach the spot according to its typology can be segmented into several regions, and consequently the "intensity" of the spot will be characterised by a vector of several parameters (*i.e.*, median and variance for each region).

This enriched set of quantified parameters (spot shape/texture features and typology, multi-region spot segmentation, multiple parameter of hybridization by spot, etc.) opens new possibilities to refine the existing microarray platforms and especially to adapt the high-level data analysis algorithms.

ACKNOWLEDGEMENTS

The author gratefully thanks Fernand Meyer for his valuable suggestions. This work is part of the French Project *GEMBIO-Bioinformatique* 2003-2006 (Mathematical methods for the analysis of biochip data: towards medical and therapeutic diagnosis and prognostic) supported by the Conseil General des Mines.

REFERENCES

Ahmed AA, Vias M, Iyer NG, Caldas C, Brenton JD (2004). Microarray segmentation methods significantly influence data precision. Nucleic Acids Res 32(5):e50.

Angulo J, Serra J (2003). Automatic analysis of DNA microarray images using mathematical morphology. Bioinformatics 19:553–62.

Angulo J (2005). Automated spot classification in cDNA images using mathematical morphology. Internal Note N-19/05/MM CMM-Ecole des Mines de Paris, 28p.

Angulo J (2007). Morphological model-based microarray spot classification and segmentation in polar coordinates. In: Proc Int Conf Stereol 2007(ICS XII), Saint-Etienne, France, September 2007, 8 p.

Appleton B, Talbot H (2005). Globally Optimal Geodesic Active Contours. J Math Imag Vision 23:67–86.

Balagurunathan Y, Dougherty ER (2002). Simulation of cDNA microarrays via a parameterized random signal model. J Biomed Opt 7: 507–23.

Bozinov D, Rahnenführer J (2002). Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering. Bioinformatics 18(5):747-56.

Brown PO, Botstein D (1999) Exploring the NewWorld of the genome with DNA microarrays. Nature Genet 21 (Suppl.):33-–7.

Chen Y, Dougherty ER, Bittner ML (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. J Biomed Opt 2: 364–74.

Demirkaya O, Asyali MH, Shoukri MM (2005). Segmentation of cDNA Microarray Spots Using Markov Random Field Modeling. Bioinformatics 21(13):2994-3000.

Ekstrom CT, Bak S, Kristensen C, Rudemo M (2004). Spot shape modelling and data transformation for microarrays. Bioinformatics 20:2270–8.

Glasbey C, Khondoker M (2005). Correction for pixel censoring in cDNA microarray. In: Proc 20th Int Worksh Stat Model, University of Western Sydney Press: 17–31.

Gottardo R, Besag J, Stephens M, Murua A (2006). Probabilistic segmentation and intensity estimation for microarray images. Biostatistics 7(1):85-99.

Lehmussola A, Ruusuvuori P, Yli-Harja O (2006). Evaluating the performance of microarray segmentation algorithms. Bioinformatics 22(23):2910–7.

Li Q, Fraley C, Bumgarner RE, Yeung KY, Raftery AE (2005). Donuts, scratches and blanks: robuts model-based segmentation of microarry images. Bioinformatics 21: 2875–82.

Luengo-Oroz MA, Angulo J, Flandrin G, Klossa J (2005). Mathematical morphology in polar-logarithmic coordinates. In: Proc 2nd Iber Conf Pattern Recogn Images Anal (IbPRIA'05), Estoril, Portugal. Springer Lect Not Comput Sci 3523:199–206.

Rosenfeld A, Pfaltz J (1968). Distance functions on digital pictures. Pattern Recogn 1:33–61.

Schena M (1968). Microarray Analysis. Hoboken, New Jersey: John Wiley and Sons.

Steinfath M, Wruck W, Seidel H, Lehrach H, Radelof U, O'Brien J (2001). Automated image analysis for array hybridization experiments. Bioinformatics 17:634–41.

Sun C, Pallottino S (2003). Circular shortest paths by branch and bound. Pattern Recogn 36:2513–20.

Vincent L (1998). Minimal Path Algorithms for the Robust Detection of Linear Features in Gray Images. In: Proc Int Symp Math Morphol (ISMM'98). Amsterdam: Kluwer, 331–8.

Wierling CK, Steinfath M, Elge T, Schulze-Kremer S, Aanstad P, Clark M, Lehrach H, Herwig R (2002). Simulation of DNA array hybridization experiments and evaluation of critical parameters during subsequent image and data analysis. BMC Bioinformatics 3:1–17.

Yang YH, Buckley MJ, Dudoit S, Speed TP (2002). Comparison of methods for image analysis on cDNA microarray data. J Comput Graph Stat 11:108–36.

APPENDIX A: COMPLEMENT ON MODEL-BASED SPOT CLASSIFICATION

The aim of this appendix is to complement the section on model-based spot classification in order to provide the definitions of the features used for the spot classification into the different typologies as well as a short analysis of performances of these features by means of a series of examples.

By means of a typical example the definition of all the descriptive parameters for $\overline{P}_\theta(\rho)$ and for $PS(n_\rho, P_\theta(\rho))$, which are the final 1D curves used to analyse the spot typology is given in Fig. 11.

The value of all these parameters has been computed for a small selection of representative spots of each typology. In Tables 2 and 3 the corresponding values are given.

Thus on the basis of these results, the following statements can be drawn. When $\sigma^{\downarrow} > 0.2$, we can suppose that a background was superposed to the spot (the degree of background is proportional to the value of $\sigma^{\downarrow}$). The parameter ρ_0 yields a rough estimate of the spot radius (very useful parameter). The spot radius can also be approached by computing $spot_{radius} = r_b^\gamma + \frac{r_{\text{spot}}^\gamma + r_b^\gamma}{2}$. The derived value $\bar{n}_{\max} = n_{\max}/\rho_0$ is associated to the spot homogeneity; typically, $\bar{n}_{\max} \leq 0.3 \leftrightarrow$ inhomogeneous spot, $0.3 < \bar{n}_{\max} < 0.6 \leftrightarrow$ homogeneous spot and $\bar{n}_{\max} > 0.6 \leftrightarrow$ very homogeneous spot. The parameter $\overline{P}_\theta(0)$ is very useful to identify some typologies. If $\overline{P}_\theta(0) \leq 0.4$ involves a doughnut or ring spot. The ambiguous situation of $0.4 < \overline{P}_\theta(0) < 0.6$ can be associated typically to a cracked spot. When $\rho_{\max} > 1$, the value of $\sigma_1 > 0.2$ allows detecting doughnut/ring spots. Then, the parameter $\rho_{\max}$ is very interesting to separate the cases doughnut/ring. The case $\rho_{\min} = 1$ is not really interesting, however when $(\rho_0 - \rho'_{\min}) > 10$, we have typically an egg or and an irregular spot. The egg-like typology is particularly associated to the case $\rho_{\max} = 1$. Concerning the significant points from $PS(P_\theta(\rho))$, and denoting $\nu_1 = \nu^{[4, r_a^\gamma]}$ and $\nu_2 = \nu^{[r_b^\gamma, r_{\text{spot}}^\gamma]}$, we have:

- If $\nu_1 < 4$ the spot is very homogeneous (regular or saturated spot).
- If $4 \leq \nu_1 \leq 15$ and $\nu_2 \geq 30$ the spot belongs typically to a cracking-like category.
- The doughnut-like and egg-like spots are associated to values of $\nu_1 > 15$ and $2 \leq \nu_2 < 30$. Ring-like spots have moreover values of $r_a^\varphi \leq 20$, and the egg-like spots have $20 < r_a^\varphi < 30$.
- Fragmented spots are characterised by very opposite values of ν_1 and ν_2, *e.g.*, $\nu_1 > 20$ and $\nu_2 < 2$ or $\nu_1 < 5$ and $\nu_2 > 30$.

APPENDIX B: IMAGE CENTROID USING GENERALIZED DISTANCE FUNCTION

Working in polar coordinates involves the selection of the center (x_c, y_c) for each spot, and this is a critical choice because if the selected center point is displaced from the "real" spot center (*i.e.*, the spot represented in polar coordinates is very "curved") it is possible that the minimal path obtained by Up/Down GMP will not be circular as well as to obtain a wrong spot classification (*e.g.*, to consider a regular spot as a fragmented spot or to miss an egg-like or a doughnut-like spot).

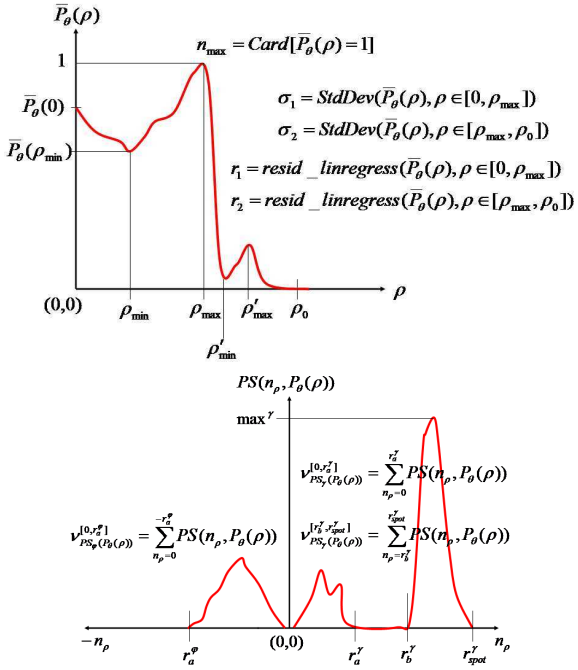


Fig. 11. *Descriptive parameters for $\overline{P}_\theta(\rho)$ (top) and for $PS(n_\rho, P_\theta(\rho))$ (bottom).*

We propose to compute the optimal (x_c, y_c) by means of the generalized distance function, reminded above in the paper. The idea is to compute in image f the GDF to the image border BB and then, to consider that the maximum of the corresponding function includes the grey-level centroid of the image. Being precise, we can proceed as follows:

- Compute GDF to set BB in image f: for each pixel p, compute $d_f(BB)$;
- Find $u_{\max}$, the maximal value of $d_f(BB)$ and threshold the result in order to keep only the pixels which values in $d_f(BB)$ are equal to $u_{\max}$: these pixels define set C.
- If C has more than one pixel, compute the centroid (using binary moments) of set C.

In Fig. 12 various examples of computation of centroid for grey-level images using this algorithm are given. As we can observe, the method is very robust and it allows detecting the optimal center for embedded structures. Note also that the GDF distance function only takes into account the bright structures. Consequently, if we are interested in the centroid of an object with a large hole, a close-holes operator can be used to compute either the centroid of the object without hole or the centroid of the hole.

Application to optimally compute spot centers in a microarray

We have used a similar algorithm based on the GDF for computing the centroid of the spots. In a microarray image, the spots are placed in blocks within an orthogonal arrangement. Using morphological operators, it is possible to build a rough estimate of the orthogonal grid of each detected block (Angulo and Serra, 2003): the block grid defines a bounding box for each spot. Then, after computing the GDF in the whole image block to the set composed by the grid, a threshold at the maximal value in each bounding box leads to the center of the corresponding spot.

In Fig. 13 an example of computation of the spot centers is depicted. Note the advantage of this approach which allows determining the optimal center even for partially overlapped spots or for spots bounded by a non precise grid.

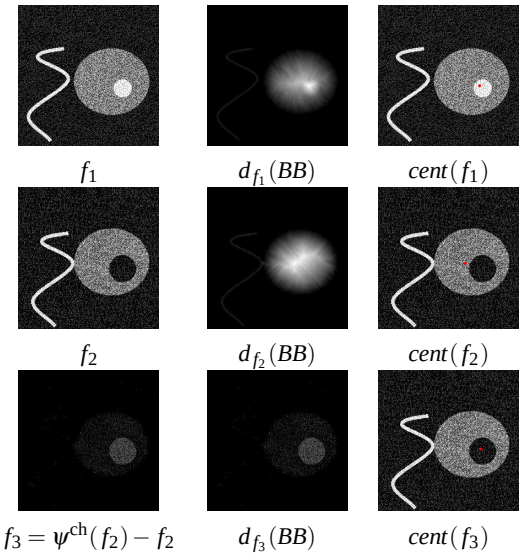


Fig. 12. *Examples of computation of image centroid using generalized distance function (GDF). The first column corresponds to the original images; second column, to the GDF to the border BB; third column, to the detected centroid (in red, and superimposed on original image). The third image corresponds to the residue of the close-holes operator.*

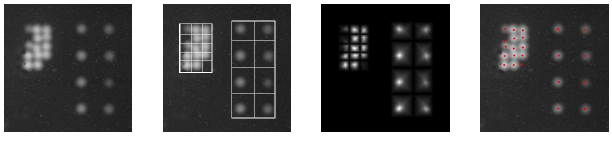


Fig. 13. *Example of computation of spot centers in a microarray image: (a) original image, (b) orthogonal grid superimposed on spots, (c) GDF of original image to grid, (d) associated center for each bounding box.*

Table 2. *Values of parameters from $\overline{P}_\theta(\rho)$ for a selection of examples of each typology.*

	Typical				Cracking			
$\sigma^{\downarrow}$	0.40	0.18	0.40	0.24	0.22	0.17	0.22	0.20
$\overline{P}_\theta(0)$	0.86	0.97	1.00	0.90	0.52	0.92	0.49	1.00
$\rho_{\max}$	55	44	1	27	48	57	63	1
ρ_0	100	74	100	56	72	80	80	78
$\rho_{\min}$	12	1	1	1	1	26	15	1
$\overline{P}_\theta(\rho_{\min})$	0.81	0.97	1.00	0.90	0.52	0.8	0.45	1.00
$n_{\max}$	9	12	51	12	13	9	5	13
σ_1	0.04	0.00	0.00	0.03	0.17	0.06	0.16	0.00
σ_2	0.41	0.38	0.42	0.38	0.38	0.38	0.38	0.27
r_1	1.76	0.10	0.00	0.45	3.49	2.66	4.57	0.00
r_2	7.06	3.29	19.8	3.21	3.59	2.52	1.35	7.44
$\rho'_{\min}$	81	73	55	55	71	79	79	26
$\overline{P}_\theta(\rho'_{\min})$	0.05	0.01	0.96	0.01	0.02	0.01	0.04	0.68
$\rho'_{\max}$	82	74	56	56	72	80	80	48
$\overline{P}_\theta(\rho'_{\max})$	0.05	0.01	0.96	0.01	0.02	0.01	0.04	0.71
	Doughnut				Egg			
$\sigma^{\downarrow}$	0.35	0.24	0.25	0.26	0.44	0.32	0.49	0.29
$\overline{P}_\theta(0)$	0.27	0.31	0.12	0.21	1.0	1.0	0.43	1.0
$\rho_{\max}$	41	57	49	39	1	1	60	1
ρ_0	100	81	75	71	75	99	100	79
$\rho_{\min}$	1	1	1	1	1	1	1	1
$\overline{P}_\theta(\rho_{\min})$	0.27	0.31	0.11	0.21	1.0	1.0	0.43	1.0
$n_{\max}$	11	8	11	9	9	7	3	17
σ_1	0.25	0.23	0.33	0.28	0.0	0.0	0.12	0
σ_2	0.44	0.38	0.37	0.37	0.35	0.35	0.34	0.39
r_1	1.87	5.17	3.56	2.36	0	0	5.72	0
r_2	7.85	2.15	2.89	2.69	7.43	14.8	7.56	11.4
$\rho'_{\min}$	56	80	74	70	11	34	79	39
$\overline{P}_\theta(\rho'_{\min})$	0.90	0.01	0.01	0.01	0.97	0.13	0.01	0.14
$\rho'_{\max}$	57	81	75	71	12	35	80	40
$\overline{P}_\theta(\rho'_{\max})$	0.90	0.01	0.01	0.01	0.97	0.13	0.01	0.14
	Fragmented				Ring		Saturated	
$\sigma^{\downarrow}$	0.29	0.27	0.21	0.19	0.45	0.26	0.20	0.18
$\overline{P}_\theta(0)$	1.0	1.0	0.97	0.91	0.15	0.18	1.0	1.0
$\rho_{\max}$	1	1	6	32	63	63	1	1
ρ_0	91	99	82	75	100	82	68	71
$\rho_{\min}$	1	1	1	1	1	1	1	1
$\overline{P}_\theta(\rho_{\min})$	1.0	1.0	0.97	0.91	0.14	0.18	1.0	1.0
$n_{\max}$	9	21	9	17	6	6	37	49
σ_1	0.0	0.0	0.01	0.03	0.22	0.26	0.0	0.0
σ_2	0.37	0.42	0.35	0.38	0.39	0.38	0.35	0.31
r_1	0	0.0	0.04	0.41	8.33	6.18	0.0	0.0
r_2	9.15	15.4	6.65	4.41	6.69	2.16	10.63	12.62
$\rho'_{\min}$	67	64	27	74	82	81	67	70
$\overline{P}_\theta(\rho'_{\min})$	0.03	0.01	0.84	0.0	0.02	0.02	0.0	0.01
$\rho'_{\max}$	68	65	28	75	83	82	68	71
$\overline{P}_\theta(\rho'_{\max})$	0.03	0.01	0.84	0.0	0.02	0.02	0.0	0.01

Table 3. *Values of parameters from $PS(n_\rho, P_\theta(\rho))$ for a selection of examples of each typology.*

	Typical				Cracking			
r_a^γ	60	52	108	24	44	60	48	56
r_b^γ	132	112	124	80	128	136	148	120
r_{spot}^γ	160	160	156	124	152	160	160	160
$v_{PS_\gamma(P_\theta(\rho))}^{[4,r_a^\gamma]}$	7.22	2.43	2.99	3.20	14.25	9.20	14.40	11.06
$v_{PS_\gamma(P_\theta(\rho))}^{[r_b^\gamma,r_{\text{spot}}^\gamma]}$	46.10	56.99	60.03	38.67	31.97	53.02	31.03	38.91
r_a^φ	24	36	32	20	8	28	28	20
$v_{PS_\varphi(P_\theta(\rho))}^{[4,r_a^\varphi]}$	1.08	1.12	0.93	0.55	0.04	3.15	1.63	1.13
	Doughnut				Egg			
r_a^γ	68	60	68	60	64	84	64	76
r_b^γ	144	148	144	128	112	132	132	124
r_{spot}^γ	160	160	160	160	148	156	156	160
$v_{PS_\gamma(P_\theta(\rho))}^{[4,r_a^\gamma]}$	32.00	19.38	31.35	25.28	19.91	22.01	15.08	21.99
$v_{PS_\gamma(P_\theta(\rho))}^{[r_b^\gamma,r_{\text{spot}}^\gamma]}$	15.15	19.85	7.91	13.26	15.25	3.71	23.25	9.48
r_a^φ	92	12	108	84	28	28	28	24
$v_{PS_\varphi(P_\theta(\rho))}^{[4,r_a^\varphi]}$	20.56	0.03	26.49	20.04	3.14	1.73	6.67	0.59
	Fragmented				Ring		Saturated	
r_a^γ	136	140	52	40	68	72	8	4
r_b^γ	148	148	72	100	152	156	52	84
r_{spot}^γ	160	160	160	160	160	160	160	160
$v_{PS_\gamma(P_\theta(\rho))}^{[4,r_a^\gamma]}$	30.28	33.61	2.78	4.11	15.55	24.12	0.03	0.03
$v_{PS_\gamma(P_\theta(\rho))}^{[r_b^\gamma,r_{\text{spot}}^\gamma]}$	0.92	0.46	44.65	50.84	5.23	11.12	51.25	58.48
r_a^φ	16	12	16	28	16	20	16	4
$v_{PS_\varphi(P_\theta(\rho))}^{[4,r_a^\varphi]}$	0.10	0.15	0.54	1.56	0.32	0.99	0.00	0.00